Hexapod Monophyly Re-Examined Based on Complete Mitochondrial Genomes.

Francesco Nardi[1]*, Giacomo Spinsanti[1], Jeffrey L. Boore[2], Antonio Carapelli[1], Romano Dallai[1] and Francesco Frati[1].

[1]Department of Evolutionary Biology, University of Siena, via Aldo Moro 2, Siena, Italy

[2]DOE Joint Genome Institute and Lawrence Berkeley National Laboratory, 2800 Mitchell Drive, Walnut Creek, CA, U.S.A.

*To whom correspondence should be addressed.

E-mail: nardifra@unisi.it

Abstract

Recent morphological and molecular evidence has changed interpretations of arthropod phylogeny and evolution. Here we compare complete mitochondrial genomes to show that Collembola, a wingless group traditionally considered as basal to all insects, appears instead to constitute a separate evolutionary lineage that branched much earlier than the separation of many crustaceans and insects, and independently adapted to life on land. Therefore, the taxon Hexapoda, as commonly defined to include all six-legged arthropods, is not monophyletic.

The phylum Arthropoda comprises the major groups Hexapoda (insects and presumed allies), Myriapoda (e.g., centipedes and millipedes), Chelicerata (e.g., spiders and horseshoe crabs), and Crustacea (e.g., crabs and lobsters). Many studies have attempted to reconstruct the evolutionary relationships among arthropods using various approaches such as palaeontology (*1*), comparative morphology (*2*), comparative developmental biology (*3, 4*) and molecular phylogenetics (*5, 6*).

It has long been held that hexapods (*7*) constitute a monophyletic taxon (*8, 9*), and that their closest relatives are to be found in myriapods (*10*). More recently, molecular and developmental studies have rejected this relationship (*3-5, 11, 12*), in favour of a closer affinity between Hexapoda and Crustacea (Pancrustacea or Tetraconata). In this context, special attention must be given to the apterygotes (springtails, silverfish and their allies), the wingless hexapods thought to branch at the base of Hexapoda. The phylogenetic position of these groups is still unclear (*13-16*), casting doubts even on the monophyly of the Hexapoda (*17*).

A potentially powerful technique for resolving deep relationships is to compare whole mitochondrial genomes (*5, 17, 18*). Phylogenetic analysis of the only complete mitochondrial sequence available for an apterygotan species (*17*) suggested the possibility that Collembola might not be included within Hexapoda, contrasting the classical view of a monophyletic taxon including all six-legged arthropods. Collembola have been clustered within crustaceans also in some other molecular and/or combined data sets (*15, 16*), but the possible paraphyly of Hexapoda has never been given specific attention and the deserved consideration. We have now sequenced the complete mitochondrial genomes of two additional species (*19*), specifically chosen to address this problem: *Tricholepidion gertschi*, representing one of the most basal lineages of the Insecta (Zygentoma), and *Gomphiocephalus hodgsoni*, another collembolan, to test support for the two competing hypotheses of a monophyletic *versus* paraphyletic Hexapoda.

An initial phylogenetic analysis performed on the 35-taxon dataset (*19*) produced the tree shown in Fig. 1. The tree has high support at most nodes, with support decreasing toward deeper relationships. This analysis strongly supports the Pancrustacea hypothesis, with the exception of the position of *Apis* and *Heterodoxus*. *Tricholepidion gertschi* is basal to all the pterygotan insects, supporting the monophyly of the Insecta. The four crustacean sequences are divided into two well defined groups (representing Malacostraca and Branchiopoda) but their reciprocal relationships and position relative to the Insecta are not resolved. The Crustacea+Insecta node is well supported, and it excludes the two collembolans, which cluster together as the basal lineage of the Pancrustacea. A second group unites the Chelicerata+Myriapoda (as in ref. # 20), but also includes the two insects *Apis* and *Heterodoxus*, presumably as an artefact.

While this tree shows many interesting outcomes, it also contains some evidently untenable relationships, which have, however, strong statistical support. This indicates the presence of anomalies in the evolution of these sequences that introduce strong systematic errors in the analysis. The most likely factors that can cause these anomalies are unequal base composition [which can bias amino acid composition (*21*)] and uneven rates of evolution among different lineages. This problem might be especially acute, since some taxa share an extremely high AT bias, *Apis* (84.8%), *Rhipicephalus* (78.0%), and *Heterodoxus* (79.3%), and different rates of evolution which could potentially cause artefactual attraction (*22*) in this analysis. Such sequences are usually removed from phylogenetic analyses due to their evidently incorrect placement and disturbance to the reconstruction. To recognise and exclude from the analysis those sequences whose placement in the phylogenetic tree could be influenced by such anomalies in the mechanism of evolution, rather than by the true historical process, we performed a detailed statistical test (*19*) to select a subset of sequences with homogeneous modes of evolution and whose rate of evolution is compatible with *Gomphiocephalus* and *Tricholepidion*. The placement of these two taxa is key to assessing the monophyly of the Hexapoda, so it is especially important that those compared be compatible with these. The methods of analysis outlined above, applied to this reduced data set, produced the two trees shown in Fig. 2, which differ only for the placement of *Ostrinia* with respect to the remaining Holometabola. Again, strong support is obtained for the Pancrustacea, with *Tricholepidion* basal to the remaining pterygotan insects, and the two collembolans placed outside the Crustacea+Insecta clade. The trees also show monophyly of Crustacea, although with lower level of support. *Limulus* is recovered as the sister group of the Pancrustacea, in contrast with the analysis based on the 35-taxon dataset, but, again, with very low support.

The most interesting result produced by this study is certainly the non-monophyly of Hexapoda, that is, the position of the two collembolans outside the Crustacea+Insecta clade, agreed upon by all analyses and with high levels of support. In order to test the relative positioning of Crustacea, Collembola, and Insecta in more detail, two alternative topologies were compared using analytical tests. The hypothesis of Crustacea external to a monophyletic Hexapoda (here, Insecta+Collembola) is strongly rejected (Tab. 1) in favour of the proposed non-monophyly of Hexapoda. The same tests were also applied to the problem of the basal trichotomy between Chelicerata, Myriapoda, and Pancrustacea. A sister group relationship between Pancrustacea and Myriapoda (=Mandibulata) is strongly rejected (Tab. 1), while no significant difference in support was found for the other two possible hypotheses. This accords with the low levels of support found in all trees at this node.

It has been generally accepted that Hexapoda, including the basal apterygotan orders, is monophyletic. This conclusion is strongly supported by similarities in their body organisation (composed of head, thorax and abdomen), as well as other morphological characters including eye and leg structure and the absence of limbs in one of the cephalic segments (*9*). On the other hand the interpretation of such characters also depends on which is the closest relative to the Hexapoda, and even on the basal splitting of these latter taxon (*9*). Nevertheless, apterygotan taxa, including Collembola, show a number of peculiar features which at least complicate the analysis of their affinities with the Insecta *sensu stricto* (*9, 23*) and leave some room to question these affinities altogether. The acceptance of non monophyly of Hexapoda implies that the tripartite and six-legged body plan typical of Hexapoda would be a convergent acquisition of collembolans and the “true insects”.

Our analysis, based on a large, specifically targeted data set and modern statistical tools, strongly supports the view that Hexapoda is not monophyletic, that at least some apterygotes have adapted to life on land independently from insects, and that those features shared between some apterygotes and insects might have originated independently in these lineages.

References and Notes

1. D. E. G. Briggs, R. A. Fortey, *Science* **246**, 241 (1989).

2. M. J. Emerson, F. R. Schram, in *Arthropod Relationships*, R. A. Fortey, R. H. Thomas, Eds. (Chapman & Hall, London, 1997), pp. 67-86.

3. C. E. Cook, M. L. Smith, M. J. Telford, A. Bastianello, M. Akam, *Curr. Biol.* **11**, 759 (2001).

4. M. Duman-Scheel, N. H. Patel, *Development* **126**, 2327 (1999).

5. J. L. Boore, D. V. Lavrov, W. M. Brown, *Nature* **392**, 667 (1998).

6. J. W. Shultz, J. C. Regier, *Proc. R. Soc. Lond. B.* **267**, 1011 (2000).

7. Hexapoda are defined as composed of the apterygotan orders Protura, Collembola, Diplura, Microcoryphia, Zygentoma, and all pterygotan orders (Pterygota). Microcoryphia + Zygentoma + Pterygota constitute the Insecta *sensu stricto*.

8. W. Hennig, *Insect phylogeny* (Wiley & Sons, New York, 1981).

9. K. -D. Klass, N. P. Kristensen, *Ann. Soc. Entomol. Fr.* **37**, 265 (2001).

10. O. Kraus, in *Arthropod Relationships*, R. A. Fortey, R. H. Thomas, Eds. (Chapman & Hall, London, 1997) pp. 295-303.

11. W. Dohle, *Ann. Soc. Entomol. Fr.* **37**, 85 (2001).

12. M. Friedrich, D. Tautz, *Nature* **376**, 165 (1995).

13. C. Bitsch, J. Bitsch, *Zoologica Scripta* **29**, 131 (2000).

14. A. Carapelli, F. Frati, F. Nardi, R. Dallai, C. Simon, *Pedobiologia* **44**, 361 (2000).

15. G. Giribet, C. Ribera, *Cladistics* **16**, 204 (2000).

16. T. Spears, G. Abele, in *Arthropod Relationships*, R. A. Fortey, R. H. Thomas, Eds. (Chapman & Hall, London, 1997) pp. 169-188.

17. F. Nardi, A. Carapelli, P. P. Fanciulli, R. Dallai, F. Frati, *Mol. Biol. Evol*. **18**, 1293 (2001).

18. E. Garcia-Machado, *et al.*, *J. Mol. Evol.* **49**, 142 (1999).

19. Materials and methods are available as supporting online material on *Science* Online.

20. U. W. Hwang, M. Friedrich, D. Tautz, C. J. Park, W. Kim, *Nature* **413**, 154 (2001).

21. K. G. Helfenbein, W. M. Brown, J. L. Boore, *Mol. Biol. Evol*. **18**, 1734 (2001).

22. D. L. Swofford, G. J. Olsen, P. J. Waddell, D. M. Hillis, in *Molecular Systematics*, D. M. Hillis, C. Moritz, B. K. Mable, Eds. (Sinauer, Sunderland, 1996) pp.407-514.

23. R. Dallai, B. A. Afzelius, in *The Male Gamete, from basic science to clinical applications*, C. Gagnon, Ed. (Cache River Press, Vienna, 1999) pp. 333-350.

24. H. Shimodaira, *Syst. Biol*. **51**, 492 (2002).

25. J. Adachi, M. Hasegawa, *Compu. Sci. Monogr.* **28**, 1 (1996).

26. J. P. Huelsenbeck, F. Ronquist, *Bioinformatics* **17**, 754 (2001).

27. This study was supported by grants from the Ministero dell'Istruzione, dell'Università e della Ricerca, from the Progetto Nazionale di Ricerca in Antartide, and from the University of Siena. Part of this work was performed under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research, by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC03-76SF00098.

Supporting Online Material

www.sciencemag.org

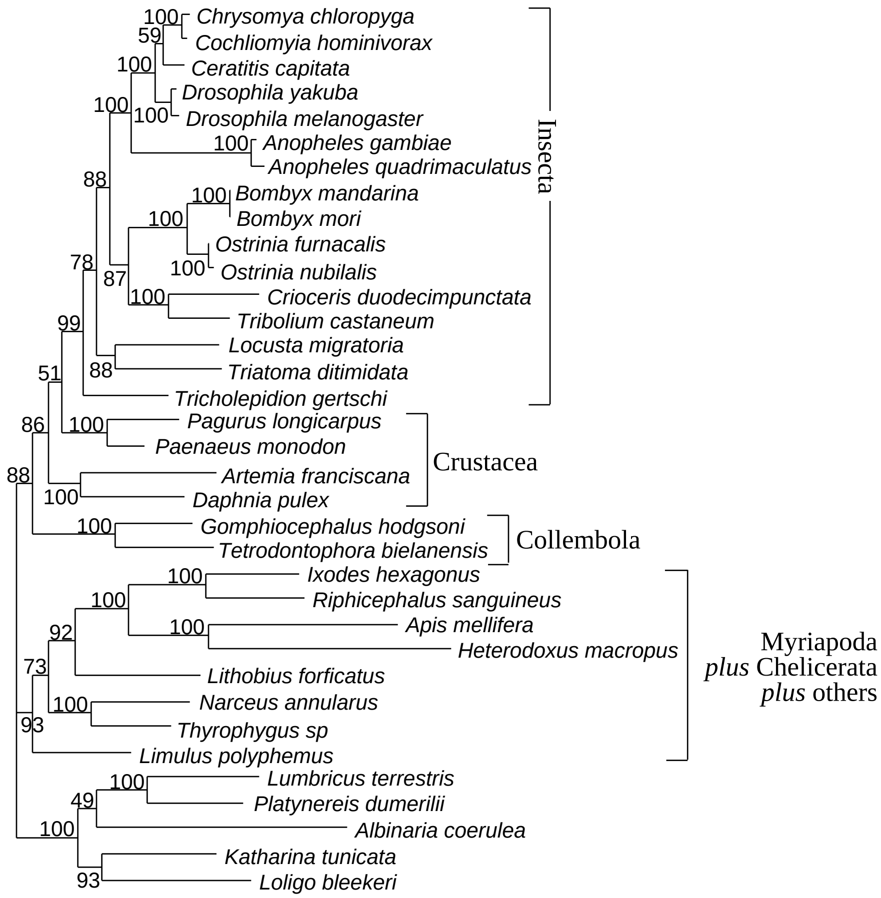
Materials and Methods

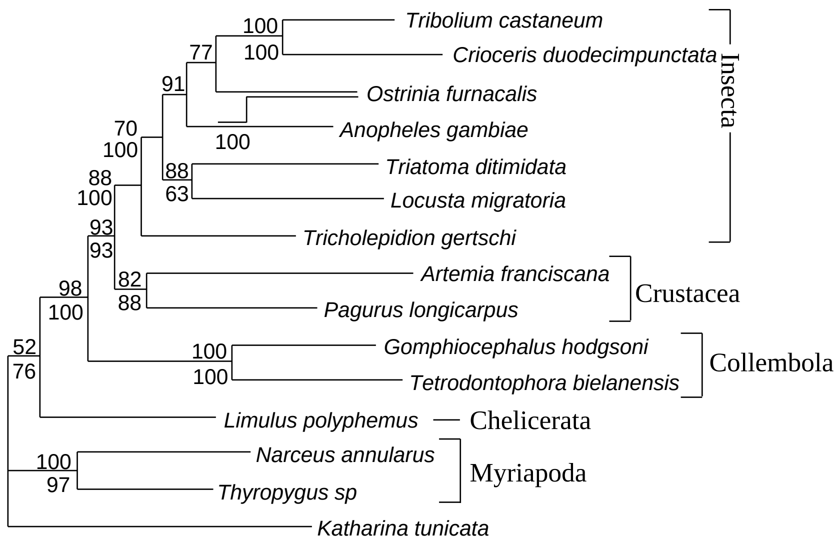
Table. 1. Tests of significance for competing hypotheses. Statistical tests of significance were conducted for different competing phylogenetic hypotheses within Pancrustacea and within arthropod classes. au: approximately unbiased test; kh: Kishino-Hasegawa test; sh: Shimodaira-Hasegawa test (*24*).

Tree	-lnL (ProtML)	au	kh	sh	
(Collembola,(Crustacea,Insecta))	19723.73	0.991	0.979	0.979	best
(Crustacea,(Collembola,Insecta))	19744.96	0.009	0.021	0.021	
(Myriapoda,(Chelicerata,Pancrustacea))	19723.73	0.509	0.496	0.649	best
((Myriapoda,Chelicerata),Pancrustacea)	19723.97	0.509	0.504	0.626	
(Chelicerata,(Myriapoda,Pancrustacea))	19739.90	0.006	0.032	0.084	

Fig. 1. Maximum Likelihood [ProtML (*25*)] phylogenetic reconstruction, complete dataset. Numerals at each node show LBP (Local Bootstrap Probability) values. Branch lengths are drawn proportionally to ML estimates.

Fig. 2. Maximum Likelihood [ProtML (*25*) and MrBayes (*26*)] phylogenetic reconstructions, reduced dataset. Alternative placement of *Ostrinia* follows MrBayes reconstruction. Numerals above each node show LBP (Local Bootstrap Probability) values (ProtML), numerals below each node indicate posterior probabilities (MrBayes). Branch lengths are drawn proportionally to ML estimates produced by ProtML.


Chrysomya chloropyga
Cochliomyia hominivorax
Ceratitis capitata
Drosophila yakuba
Drosophila melanogaster
Anopheles gambiae
Anopheles quadrimaculatus
Bombyx mandarina
Bombyx mori
Ostrinia furnacalis
Ostrinia nubilalis
Crioceris duodecimpunctata
Tribolium castaneum
Locusta migratoria
Triatoma ditimidata
Tricholepidion gertschi
Pagurus longicarpus
Paenaeus monodon
Artemia franciscana
Daphnia pulex
Gomphiocephalus hodgsoni
Tetrodontophora bielanensis
Ixodes hexagonus
Riphicephalus sanguineus
Apis mellifera
Heterodoxus macropus
Lithobius forficatus
Narceus annularus
Thyrophygus sp
Limulus polyphemus
Lumbricus terrestris
Platynereis dumerilii
Albinaria coerulea
Katharina tunicata
Loligo bleekeri
Insecta
Crustacea
Collembola
Myriapoda
plus Chelicerata
plus others


Tribolium castaneum
Crioceris duodecimpunctata
Ostrinia furnacalis
Anopheles gambiae
Triatoma ditimidata
Locusta migratoria
Tricholepidion gertschi
Insecta
Artemia franciscana
Pagurus longicarpus
Crustacea
Gomphiocephalus hodgsoni
Tetrodontophora bielanensis
Collembola
Limulus polyphemus
Chelicerata
Narceus annularus
Thyropygus sp
Myriapoda
Katharina tunicata
100
100
77
91
100
70
100
88
100
88
63
93
93
82
88
98
100
100
100
52
76
100
97